DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Field of the Invention]A dictionary preparation device for speech recognition and a dictionary preparation method for speech recognition which draw up the dictionary used for the voice recognition equipment [invention / this] for an unspecified speaker, It is related with the voice recognition equipment using the drawn-up dictionary, the personal digital assistant machine carrying this voice recognition equipment, and the program recording medium that recorded the dictionary creation processing program.

[0002]

[Description of the Prior Art]There is a method of registering reading acquired by conducting the morphological analysis of the above-mentioned character string information, and searching for reading of the notation as a method of conventionally drawing up the dictionary for speech recognition from the character string information containing a sentence mixing kanji, kana and characters. In a web browser, although there is application which displays a web page based on a speech recognition result, when calling the above-mentioned web page to the dictionary for speech recognition used in that case and registering beforehand the word which a user utters, the above-mentioned dictionary preparation method is applied. For example, by relating with URL (uniform resource location) of the Prime Minister's official residence the word the „Prime Minister's official residence", and memorizing it, if the „Prime Minister's official residence" and a user utter, the homepage of the Prime Minister's official residence can be displayed.This is realized by registering with the dictionary for speech recognition in quest of reading of „******************" from the notation the „Prime Minister's official residence."

[0003]Adjacency relation is extracted from a lot of text databases for study, and there is the method of drawing up the dictionary for speech recognition (language model) in quest of statistical connection information. for example, the statistical chain relation between each word

which divided character string information into each word (morpheme) automatically, and was divided in the voice recognition equipment indicated to JP,11-259088,A using the morphological-analysis program etc. -- a bigram, a trigram, etc. are specifically calculated. It is a Chinese character of the same notation, and when there is two or more reading, it asks for the frequency of two or more reading of each. And an input of a sound will compute language likelihood at each word which constitutes each recognition sentence candidate's word series combining the probability value of the above-mentioned statistical chain relation, and the frequency of reading to each recognition sentence candidate's likelihood which generates a recognition sentence candidate and is calculated from a language model using the extracted feature parameter. And the recognition result has been obtained based on this language likelihood.

[0004]

[Problem(s) to be Solved by the Invention]However, there are the following problems in the above-mentioned conventional dictionary preparation method for speech recognition. That is, in the case of the dictionary preparation method which registers first reading of the notation obtained from the morphological analysis of the above-mentioned character string information, there is a problem that the utterance unit showing character string information and the connection word uttered is not necessarily in agreement. The case where a homepage is hereafter displayed by a browser using the speech recognition result mentioned above is explained to an example. There are URL and a title in the information on the above-mentioned homepage. URL is the information showing the whereabouts of a homepage and is written like ·http://www.kantei.go.jp.·A title is a title of the page displayed on the position decided by the browser, and are text, such as Chinese character writing both in kanji and kana like a ·Prime Minister's official residence top page·, and the alphabet.

[0005]When applying the dictionary preparation method concerned to the above-mentioned title ·Prime Minister's official residence top page· and generating a recognized vocabulary automatically from the above-mentioned title, the morphological analysis of a title ·Prime Minister's official residence top page· is conducted, and it can realize by specifying reading of a word ·Prime Minister's official residence top page.·

[0006]However, when the dictionary preparation method concerned is applied. In order to read the title ·Prime Minister's official residence top page· as it is and to only change and register it into ·****************** convex ****-**·, a recognized vocabulary serves as a ·Prime Minister's official residence top page·, The partial character strings ·Prime Minister's official residence· of the above-mentioned title ·Prime Minister's official residence top page· cannot be started, and it cannot be made a recognized vocabulary. Therefore, when the partial utterance the ·Prime Minister's official residence· is made, for example, voice inputting ·Prime Minister's official residence· can be recognized using the dictionary for speech recognition drawn up as

mentioned above.

[0007]Since neither all the combination of partial character strings nor the utterance probability between partial character strings is taken into consideration even though the partial character strings which serve as a keyword from the title of the above-mentioned homepage are started temporarily and it creates a recognized vocabulary, there is a problem that the practical dictionary for speech recognition cannot be drawn up.

[0008]Next, in the case of the dictionary preparation method which draws up the dictionary for speech recognition in quest of the above-mentioned statistical connection information, the above problems are not generated in order to recognize voice inputting using the pause statistical adjacency between each word. The more practical dictionary for speech recognition may be able to be drawn up in that a recognition sentence candidate's appearing probability also including the frequency of each reading of a word with two or more reading is computed.

[0009]However, only by using appearing probability, Consideration that the probability that former one will be uttered with the notation of the front portion of a title and the notation of a trailer is high (for example, the probability that the "Prime Minister's official residence" will be uttered from a "top page" is high in the case of a title "top page of the Prime Minister's official residence") is not carried out, There is a problem that the probability that a desired recognition result will be obtained is low.

[0010]It is thought that they are not uttered if a "homepage" and specific vocabularies, such as "Welcome!", are independent. However, since consideration to it is not carried out, either, erroneous recognition of another utterance may be carried out to the vocabulary a "homepage."

[0011]The dictionary preparation method which draws up the dictionary for speech recognition (language model) from such a lot of study text databases, Although it is good for the use which recognizes a vocabulary of tens of thousands of words like dictation, there is a problem of the recognition performance that it is difficult the problem of the cost that high-speed CPU (central processing unit) and a lot of storage capacities are required, and to find out a right thing out of many vocabularies dramatically.

[0012]Then, a dictionary preparation device for speech recognition and a dictionary preparation method for speech recognition with which, as for the purpose of this invention, high recognition precision is acquired by low cost, It is in providing the voice recognition equipment using the drawn-up dictionary, the personal digital assistant machine carrying this voice recognition equipment, and the program recording medium that recorded the dictionary creation processing program.

[0013]
[Means for Solving the Problem]In order to attain the above-mentioned purpose, a dictionary preparation device for speech recognition of the 1st invention, An analysis means to analyze

inputted character string information, to divide into a composition word, and to output all the division candidates, A reading grant means to give reading to each composition word divided [above-mentioned], and to output all the reading candidates, By all the division candidates and the above-mentioned reading grant means which were obtained by the above-mentioned analysis means. A lexical preparing means which generates several utterance units from which one utterance unit which changes with one word or a connection word to which reading was given based on all the obtained reading candidates, or combination of a word differs as a recognized vocabulary, It is characterized by having a lexical memory measure which memorizes each recognized vocabulary generated [above-mentioned] as a dictionary for speech recognition.

[0014]According to the above-mentioned composition, based on all the division candidates obtained from one character string information, and all the reading candidates, one or more utterance units to which reading was given by lexical preparing means are generated. Therefore, a dictionary for speech recognition which makes an utterance unit with the possibility of utterance a recognized vocabulary from given character string information is generated by registering an utterance unit generated in this way as a recognized vocabulary. That is, even if a user utters which partial character strings in a character string set up beforehand, a dictionary for speech recognition which can realize voice recognition equipment which can be recognized correctly is drawn up.

[0015]A dictionary preparation device for speech recognition of an invention of the above 1st, It accomplishes so that analysis likelihood which expresses a probability as an analysis result of the above-mentioned input string to each analytical candidate who becomes in a sequence of a composition word divided [above-mentioned] about the above-mentioned analysis means may be given, It accomplishes so that reading likelihood showing a probability as reading of the above-mentioned input string may be given to a sequence of reading given to a word which constitutes each above-mentioned analytical candidate in the above-mentioned reading grant means, By the above-mentioned lexical preparing means. The word order of appearance showing the order of appearance in the above-mentioned input string of a head word which constitutes the above-mentioned analysis likelihood of an analytical candidate in whom each generated utterance unit exists, the above-mentioned reading likelihood of an analytical candidate in whom each above-mentioned utterance unit exists, and each above-mentioned utterance unit, the number of Maura of each above-mentioned utterance unit, and each above-mentioned utterance unit. It has an utterance probability calculating means which calculates utterance probability of each utterance unit generated [above-mentioned] using at least one of the word frequency of occurrence showing the frequency of occurrence of a word which appears the fewest in all the character string information inputted among words to constitute, and the key word dictionary collated results, It is desirable to give utterance probability

computed [above-mentioned] in a recognized vocabulary which becomes in each above-mentioned utterance unit about the above-mentioned lexical preparing means, and to accomplish so that the above-mentioned lexical memory measure may be made to memorize.

[0016]According to the above-mentioned composition, utterance probability computed by having used at least one of analysis likelihood, reading likelihood, the word order of appearance, the number of Maura, the word frequency of occurrence, and the key word dictionary collated results for a recognized vocabulary which becomes in each above-mentioned utterance unit is given and registered into the above-mentioned lexical memory measure. Therefore, even if a recognized vocabulary generated as a result of incorrect analysis by the above-mentioned analysis means and a recognized vocabulary which is not uttered are registered into the above-mentioned dictionary for speech recognition, Utterance probability of such an unnecessary recognized vocabulary is set up small, and a dictionary for speech recognition which can realize voice recognition equipment which presents high recognition precision is drawn up.

[0017]A dictionary preparation device for speech recognition of an invention of the above 1st, An incorporation means to incorporate contents including character string information, and an extraction condition storing means in which an extraction condition for extracting character string information required for dictionary creation was stored, It is desirable to have a character-string-information extraction means to extract character string information required for dictionary creation with reference to the above-mentioned extraction condition out of character string information in contents which were incorporated as for the account of the upper, and to send out to the above-mentioned analysis means.

[0018]According to the above-mentioned composition, the above-mentioned dictionary for speech recognition is automatically drawn up from contents information by storing an extraction condition which used the feature of contents for an extraction condition storing means.

[0019]As for a dictionary preparation device for speech recognition of an invention of the above 1st, it is desirable to accomplish so that information on a web page currently shown by web browser considering the above-mentioned incorporation means as the above-mentioned contents may be incorporated.

[0020]According to the above-mentioned composition, by storing in the above-mentioned extraction condition storing means "when a <title> tag exists, a character string surrounded by <title> and </title> is extracted", for example, A character string "top page of the Prime Minister's official residence" is extracted from a top page </title> of the title "<title> Prime Minister's official residence of a web page."And based on the above-mentioned character string "top page of the Prime Minister's official residence", a dictionary for speech recognition is drawn up automatically as mentioned above.

[0021]As for a dictionary preparation device for speech recognition of an invention of the above 1st, it is desirable to accomplish so that information on a TV program text-ized considering the above-mentioned picking ****** as the above-mentioned contents may be incorporated.

[0022]a character string which corresponds to a tag "program name" by storing in the above-mentioned extraction condition storing means "a character string with a tag called a program name is extracted", for example according to the above-mentioned composition -- "-- NHK news -- good morning, Japanese" is extracted. and the above-mentioned character string -- "-- NHK news -- good morning, based on Japanese", a dictionary for speech recognition is drawn up automatically as mentioned above.

[0023]A dictionary preparation device for speech recognition of an invention of the above 1st, It has a similarity calculation means which calculates acoustical similarity between each utterance unit generated by the above-mentioned lexical preparing means, and, as for the above-mentioned lexical preparing means, it is desirable to accomplish so that utterance probability given to each above-mentioned recognized vocabulary may be changed according to similarity computed [above-mentioned].

[0024]When according to the above-mentioned composition similarity of an utterance unit "prime minister" and an utterance unit "opinion" which were generated by the above-mentioned lexical preparing means is higher than a predetermined value and both utterance units are acoustically similar, For example, while a value of utterance probability of an utterance unit "prime minister" that a value of utterance probability will play a central role in an input string highly is raised further, a value of utterance probability of an utterance unit "opinion" which is not so can lower further. By carrying out like this, not a homepage of "a top page of the Prime Minister's official residence" which erroneous recognition of the utterance "prime minister" which plays a central role is carried out to an "opinion", for example, is made into the purpose but a homepage of "an opinion of the Department of Justice" is prevented from being displayed.

[0025]A dictionary preparation method for speech recognition of the 2nd invention, A step which analyzes inputted character string information, divides into a composition word, and outputs all the division candidates, A step which gives reading to each composition word divided [above-mentioned], and outputs all the reading candidates, A step which generates several utterance units from which one utterance unit which changes with one word or a connection word to which reading was given based on all the reading candidates obtained as a result of all the division candidates and the above-mentioned reading grant which were obtained as a result of the above-mentioned word division, or combination of a word differs as a recognized vocabulary, It is characterized by having a step which memorizes each recognized vocabulary generated [above-mentioned] as a dictionary for speech recognition.

[0026]According to the above-mentioned composition, one or more utterance units to which reading was given as well as a case of an invention of the above 1st are generated. Therefore,

a dictionary for speech recognition which makes an utterance unit with the possibility of utterance a recognized vocabulary from given character string information is generated by registering an utterance unit generated in this way as a recognized vocabulary. That is, even if it utters partial character strings of a character string set up beforehand, a dictionary for speech recognition which can realize voice recognition equipment which can be recognized correctly is drawn up.

[0027]The 3rd invention is voice recognition equipment which recognizes by performing collation with a recognized vocabulary registered into a dictionary in an inputted sound, and is characterized by using a dictionary for speech recognition drawn up by a dictionary preparation device for speech recognition of an invention of the above 1st as the above-mentioned dictionary.

[0028]According to the above-mentioned composition, collation with a dictionary for speech recognition which makes a recognized vocabulary an utterance unit with the possibility of utterance generated from given character string information is performed, and voice inputting is recognized. Therefore, even if it utters partial character strings of a character string set up beforehand, it is recognized correctly.

[0029]Voice recognition equipment of an invention of the above 3rd accomplishes the above-mentioned dictionary with a dictionary for speech recognition drawn up by a dictionary preparation device for speech recognition which incorporates information on a web page as the above-mentioned contents, It is desirable to have a web page displaying means which displays a web page according to a recognition result, and a control means which switches and controls display information of the above-mentioned web page displaying means based on the above-mentioned recognition result.

[0030]Since a dictionary for speech recognition automatically drawn up from information on the above-mentioned web page is used according to the above-mentioned composition, a title of a web page, etc. are recognized correctly. Therefore, a web page according to a recognition result is correctly displayed on the above-mentioned web page displaying means by display information of a web page displaying means switching, and controlling it by a control means based on a recognition result.

[0031]Voice recognition equipment of an invention of the above 3rd accomplishes the above-mentioned dictionary with a dictionary for speech recognition drawn up by a dictionary preparation device for speech recognition which incorporates information on a TV program as the above-mentioned contents, A television displaying means which displays a TV program according to a recognition result, and a recording means which records a TV program according to a recognition result, It is desirable to have a reproduction means which plays a TV program recorded by the above-mentioned recording means, and a control means which controls the above-mentioned television displaying means and recording means, and a

reproduction means, and performs setting out of a change and a recording condition of a display channel or playback of a picture recording program based on the above-mentioned recognition result.

[0032]Since a dictionary for speech recognition automatically drawn up from information on the above-mentioned TV program is used according to the above-mentioned composition, a TV program name etc. are recognized correctly. Therefore, a change of a display channel, setting out of a recording condition, or reproduction of a picture recording program is correctly performed by controlling a television displaying means, a recording means, and a reproduction means by a control means based on a recognition result.

[0033]Voice recognition equipment of an invention of the above 3rd by a collation means which performs collation with an auxiliary dictionary in which a recognized vocabulary obtained without being based on an analysis result of specific character string information was registered, and the above-mentioned dictionary and an auxiliary dictionary and the above-mentioned collation means. A search means to search a character string which corresponds to the recognition result concerned out of text relevant to character string information inputted into the above-mentioned dictionary preparation device for speech recognition when a recognized vocabulary registered into the above-mentioned auxiliary dictionary as the above-mentioned recognition result is chosen, and drawing up the above-mentioned dictionary, It is desirable to have a selecting means which chooses a character string registered into the above-mentioned dictionary from two or more character strings searched [above-mentioned].

[0034]Even when a vocabulary which is not registered into a dictionary drawn up by a dictionary preparation device for speech recognition of an invention of the above 1st is uttered according to the above-mentioned composition, the vocabulary is recognized correctly. When not a recognized vocabulary registered into the above-mentioned dictionary but a recognized vocabulary registered into an auxiliary dictionary drawn up without being based on a dictionary preparation device for speech recognition of an invention of the above 1st is chosen as a recognition result, by a search means. For example, it was inputted into the above-mentioned dictionary preparation device for speech recognition, a character string applicable to the recognition result concerned is searched out of web page information relevant to a title of a web page. And a character string registered into the above-mentioned dictionary is chosen by selecting means from two or more character strings searched [above-mentioned]. Therefore, by registering the vocabulary into the above-mentioned dictionary, the number of recognized vocabularies of the above-mentioned dictionary increases, and recognition speed improves.

[0035]The 4th invention is voice recognition equipment which recognizes by performing collation with a recognized vocabulary registered into a dictionary in an inputted sound, A dictionary preparation device for speech recognition of an invention of the above 1st is carried, and it is characterized by using a dictionary for speech recognition drawn up by the above-

mentioned dictionary preparation device for speech recognition as the above-mentioned dictionary.

[0036]According to the above-mentioned composition, by inputting character string information into a dictionary preparation device for speech recognition carried, an utterance unit which has the possibility of utterance from this character string information is generated, and a dictionary for speech recognition which makes this utterance unit a recognized vocabulary is drawn up. Therefore, by performing collation with this dictionary for speech recognition, and recognizing voice inputting, even if it utters partial character strings of a character string set up beforehand, it is recognized correctly.

[0037]A personal digital assistant machine of the 5th invention is characterized by carrying voice recognition equipment of the above 3rd and the 4th invention.

[0038]It is better for operativity to be based on utterance rather than key operation in a personal digital assistant machine, when performing operator guidance. According to the above-mentioned composition, even if it utters which partial character strings in a character string set up beforehand, voice recognition equipment using a dictionary for speech recognition which can be recognized correctly is carried. Therefore, even if it does not utter correctly wording for performing operator guidance in a destination etc. as decided beforehand, operation of a call of a homepage, etc. is performed correctly, for example.

[0039]A program recording medium of the 6th invention is characterized by recording a dictionary creation processing program as which a computer is operated as an analysis means in the 1st above-mentioned invention, a reading grant means, a lexical preparing means, and a lexical memory measure.

[0040]According to the above-mentioned composition, one or more utterance units to which reading was given as well as a case of an invention of the above 1st are generated. Therefore, by registering an utterance unit generated in this way as a recognized vocabulary, even if it utters partial character strings of a character string set up beforehand, a dictionary for speech recognition which can realize voice recognition equipment which can be recognized correctly is drawn up.

[0041]

[Embodiment of the Invention]Hereafter, the embodiment of a graphic display of this invention explains in detail.

<1st embodiment> drawing 1 is a block diagram in the dictionary preparation device for speech recognition of this embodiment. If character string information is inputted into the analyzing processing part 1, the language of an input string will be analyzed by the text analyzing part 2, and it will be divided into a morpheme. When two or more division candidates exist in that case, all the division candidate is outputted. And reading of the morpheme divided [above-mentioned] is given by the reading grant part 3. When two or more reading exists in that case,

all the reading is outputted. When the above-mentioned text analyzing part 2 conducts text analysis, the language data containing a required analysis dictionary are stored in the analysis dictionary memory 4.

[0042]The lexical preparing part 5 draws up the dictionary for speech recognition required in order to read with the text-analysis result by the above-mentioned text analyzing part 2 and to perform speech recognition based on the reading grant result by the grant part 3. The vocabulary storage section 6 memorizes the dictionary for speech recognition drawn up by the lexical preparing part 5. And this dictionary for speech recognition is used at the time of speech recognition.

[0043]Drawing 2 is a flow chart of the dictionary creation processing operation performed by each part of the dictionary preparation device for speech recognition which has the above-mentioned composition. Hereafter, according to the flow chart of drawing 2, operation of the dictionary preparation device for real-intention voice recognition is explained to an example for the case where the character string information of "Okubo, Urawa" is inputted into the analyzing processing part 1. Here the input of the character string information (text) over the text analyzing part 2, May be an input by the inputs and the received teletexts from a network, such as WWW (World Wide Web), and, The input from sentence input means, such as a keyboard and a pen, or the input of the recognition result from voice recognition equipment may be used, and it does not matter even if it is an input from character readers, such as OCR (optical character reader).

[0044]Character string information "Okubo, Urawa" is incorporated by the above-mentioned text analyzing part 2 at Step S1. It is Step S2 and text analysis into which the input string "Okubo, Urawa" is further divided at a morpheme (word) with reference to the analysis dictionary stored in the analysis dictionary memory 4 is conducted. two [as a result,], a word sequence "Urawa (noun)", "a city (suffix)", "Okubo (noun)", and a word sequence "Urawa (noun)", "city size (noun)" and "Kubo (noun)", -- it divides and a candidate is obtained. An example of the output of the text analyzing part 2 is shown in drawing 3.

[0045]At Step S3, reading is given to each word by the above-mentioned reading grant part 3, and a reading grant result as shown in drawing 4 is obtained. Here, reading of each word is given with reference to the analysis dictionary stored in the analysis dictionary memory 4. All the reading is given to the word in which two or more reading exists. In drawing 4, it reads to a character "city", and reads to it, saying "carry out", and two reading with "****" is given to it. Hereafter, processing shifts to the lexical preparing part 5.

[0046]By step S4, it obtained and divides by the above-mentioned text analyzing part 2, and the number of candidates is set to the variable k. Since two kinds are divided as "Urawa / city / Okubo", and "Urawa / city size / Kubo" and a candidate exists as shown in drawing 3 in the case of the above-mentioned input string "Okubo, Urawa", "2" is set to the variable k. At Step S5, it

divides and the initial value ·1· is set to the candidate number m. at Step S6, the m-th total number of words that divides and constitutes a candidate is strange -- it is set to several Nm. the case of the above-mentioned input string ·Okubo, Urawa· -- the 1st -- since it divides and the candidates ·Urawa / city / Okubo· are three words of ·Urawa (noun)·, ·a city (suffix)·, and ·Okubo (noun)·, ·3· is set to the variable N1.

[0047]The number j of words after the second which constitutes the utterance unit registered as a vocabulary from Step S7 is initialized by ·0··At Step S8, the position (number of a head word) i of an utterance unit is initialized by ·1··By step S9, it is distinguished whether the value of (i+j) is the total word of several Nm or less. As a result, in being the total word of several Nm or less, while progressing to Step S10, in being larger than the total several Nm word, it progresses to Step S13. The word "Wi" which constitutes an utterance unit from Step S10, --, the reading Yi corresponding, with Wi+j (1), --, Yi (1+pi), --, Yi+j (1), --, Yi+j (1+pi+j) (however, pi+j: the number of reading after the second of the words Wi+j) are registered into the vocabulary storage section 6. Therefore, when there is two or more reading over one word "Wi", like ·Yi (1), Yi (2) --·, all the reading is registered.

[0048]At Step S11, it **************s the value of the position i of the above-mentioned utterance unit. At Step S12, it is distinguished whether the position i of an utterance unit is the total word of several Nm or less. As a result, in being the total word of several Nm or less, it shifts to registration of the utterance unit which returns to step S9 and is in the position of the following. On the other hand, in being larger than the total several Nm word, it progresses to Step S13. It **************s the number j of words after the second which constitutes an utterance unit from Step S13. At Step S14, it is distinguished whether the above-mentioned number j of words is smaller than the total several Nm word. As a result, in being smaller than the total several Nm word, it returns to Step S8 and shifts to registration of an utterance unit with many one words after the second. On the other hand, in being a not less than several Nm word, it progresses to Step S15. At Step S15, it divides and **************s the candidate number m. At Step S16, it divides, the candidate number m divides, and it is distinguished whether it is several k or less [of a candidate]. As a result, it divides, and in being below the number k of candidates, it returns to Step S6, the next divides, and it shifts to the processing to a candidate. On the other hand, it divides, it is judged that all the processings divide and concerning a candidate were completed when larger than the number k of candidates, and dictionary creation processing operation is ended.

[0049]When the above dictionary creation processing operation is performed to the above-mentioned character string information ·Okubo, Urawa· as for the above result, registration is performed as follows to the vocabulary storage section 6. That is, when [1st] it divides and the number j of words after the second of utterance units is ·0· first to a candidate ·Urawa / city / Okubo·, the pair ·Urawa/******· of an utterance unit / reading, ·city/Carrying out·, and ·Okubo / **

to set· is registered into small order for the position i of an utterance unit. Next, when the above-mentioned number j of words of an utterance unit is ·1·, the pair ·Urawa/****** carries out· of an utterance unit / reading, and ·city Okubo / ** which is carried out and is set· is registered into small order for the position i of an utterance unit. next -- the case where the above-mentioned number j of words of an utterance unit is ·2· -- the pair of an utterance unit / reading -- ·-- Urawa Okubo / reverse side -- ****** -- **· to set set is registered.

[0050]When [2nd] it divides and the above-mentioned number j of words of an utterance unit is ·0· first to a candidate ·Urawa / city size / Kubo·, ·immediately after the pair ·Urawa/******· of an utterance unit / reading, and city size/·, ·Kubo/****· is registered into small order for the position i of an utterance unit. that time -- the pair of other utterance unit / reading -- ·-- they are city size / **** -- it is -- ·-- it is registered. next -- the case where the above-mentioned number j of words of an utterance unit is ·1· -- the pair ·Urawa large / reverse side ********· of an utterance unit / reading, ·city Okubo / order ****·, and the pair of other utterance unit / reading -- ·-- they are the Urawa large / reverse side ****** -- it is -- ·-- ·-- they are city Okubo / **** -- **· which goes is registered. next -- the case where the above-mentioned number j of words of an utterance unit is ·2· -- the pair ·Urawa Okubo / reverse side ************· of an utterance word / reading, and the pair of other utterance unit / reading -- ·-- they are Urawa Okubo / reverse side ****** -- **· which goes is registered.

[0051]As a result, as shown in <u>drawing 5</u>, a recognized vocabulary will be registered into the above-mentioned vocabulary storage section 6, and the dictionary for speech recognition will be drawn up.

[0052]As mentioned above, in this embodiment, by the text analyzing part 2, an input string is divided into a morpheme (word), all divide, and a candidate is called for. All divide and reading is given by the reading grant part 3 to a candidate's word. In that case, all the reading is given to the word in which two or more reading exists. And all divide, and he creates the pair of an utterance unit / reading in consideration of the combination of a candidate, all the reading candidates, and all the connection words, and is trying to register with the vocabulary storage section 6 by the lexical preparing part 5.

[0053]That is, according to this embodiment, the vocabulary of two or more utterance units is generable from input string information. Therefore, it becomes possible from one input string information to draw up the dictionary for speech recognition which can be recognized even if it utters which partial character strings in the character string concerned.

[0054]For example, by inputting the title of a homepage into the text analyzing part 2, and drawing up the dictionary for speech recognition by an above-mentioned procedure, when calling a homepage with a sound on a web, The homepage can be called even if the above-mentioned title and utterance are not thoroughly in agreement. For example, when the title of a homepage is ·a top page of the Prime Minister's official residence·, even if it utters with ·the top

page of the Prime Minister's official residence" and utters with the "Prime Minister's official residence", Or even if it utters with "the top page of an official residence", the dictionary for speech recognition which can call the homepage of the Prime Minister's official residence can be obtained.

[0055]By similarly, inputting the program name of an electronic television program listing into the text analyzing part 2, and drawing up the dictionary for speech recognition by an above-mentioned procedure, when utterance of a TV program name performs transfer in channel of television, for example, a program name -- "-- NHK news good morning -- when it is Japanese", even if it utters with "NHK news" -- "-- good morning, even if it utters with Japanese", if predetermined time comes, the dictionary for speech recognition which can switch the channel of television to "NHK" automatically can be obtained.

[0056]in order to explain simply -- a character string "NHK news" -- "-- good morning, the case where Japanese" is registered into the dictionary for speech recognition as one word is explained to the example. However, the dictionary creation for continuous speech recognitions is also realizable in a similar way by remembering "news", and "good morning" and "Japan" as an independent word, and memorizing separately the information that these words continue. ["NHK", and]

[0057]According to a 1st embodiment of the <2nd embodiment> above, the utterance unit of only the combination of the connection word obtained as a result of text analysis is acquired, but incorrect analysis is also included in these utterance units. Therefore, treating all the acquired utterance units equally can become the cause of reducing a recognition rate, including an unnecessary vocabulary mostly. In such a case, this embodiment copes with it.

[0058]Drawing 6 is a block diagram in the dictionary preparation device for speech recognition of this embodiment. An input string will be divided into a morpheme by the text analyzing part 12 if character string information is inputted into the analyzing processing part 11. In that case, when two or more division candidates exist, the analysis likelihood which expresses the degree of a probability to all the division candidate is given and outputted. And reading of the morpheme divided [above-mentioned] is given by the reading grant part 13. In that case, when two or more reading exists, the reading likelihood showing the degree of a probability is given and outputted to all the reading. When giving by the time of the text analyzing part 12 conducting text analysis and the reading grant part 13 reading like the case of a 1st embodiment of the above, the language data containing a required analysis dictionary are stored in the analysis dictionary memory 14.

[0059]The lexical preparing part 15 draws up the dictionary for speech recognition required in order to perform speech recognition based on the text-analysis result by the text analyzing part 12, the reading grant result by the reading grant part 13, and the utterance probability by the utterance probability calculation part 16. The analysis likelihood to which the text analyzing

part 12 outputs the utterance probability calculation part 16, the reading likelihood which the reading grant part 13 outputs, The utterance probability of the utterance unit acquired from each part or ****** by the lexical preparing part 15 is computed using at least one of the word order of appearance, the number of Maura, the word frequency of occurrence, and the key word dictionary collation. The vocabulary storage section 17 memorizes the dictionary for speech recognition drawn up by the lexical preparing part 15. And this dictionary for speech recognition is used at the time of speech recognition.

[0060]An example of the output of the text analyzing part 12 is shown in drawing 7. An example of the output of the reading grant part 13 is shown in drawing 8. An example of the output of the lexical preparing part 15 is shown in drawing 9.

[0061]Hereafter, operation of the utterance probability calculation part 16 which is the feature in this embodiment is described in detail. In this utterance probability calculation part 16, as mentioned above, the utterance probability of each utterance unit is computed using at least one of the analysis likelihood KS by the text analyzing part 12, the reading likelihood YS by the reading grant part 13, the word order of appearance, the number of Maura, the word frequency of occurrence, and the key word dictionary collation.

[0062]First, the case where it asks for the above-mentioned utterance probability from the analysis likelihood KS obtained by the above-mentioned text analyzing part 12 is explained. The above-mentioned analysis likelihood KS becomes an index which plans how probable the result (morphological division results) obtained by analyzing an input string is. The same input string as the case of a 1st embodiment of the above "Okubo, Urawa" is raised to an example, and is explained.

[0063]Suppose that the 1st analytical candidate's (dividing candidate) "Urawa / city / Okubo" analysis likelihood was obtained with KS (1) by the above-mentioned text analyzing part 12, and the 2nd analytical candidate's "Urawa / city size / Kubo" analysis likelihood was obtained with KS (2). The number of the utterance units acquired combining the composition word contained in the i-th analytical candidate is made into N(i). However, when the same utterance unit is included in two or more analytical candidates, suppose that the utterance unit of an analytical candidate with the highest analysis likelihood is counted. If it is a case of the above-mentioned input string "Okubo, Urawa", since the utterance unit which a composition word comprises in "Urawa" is included in the 1st and both the 2nd analytical candidates, it will be counted only at analytical candidates with high analysis likelihood.

[0064]A total combination (that is, all the utterance units) of the composition word in one analytical candidate shall be generated in equivalent probability. If it does so, the utterance probability P1 of the utterance unit w in the i-th analytical candidate (w) can be expressed by a formula (1). However, M in a formula (1) is the number of analytical candidates.

$$P1(w) = KS(i) / \sum_{j=1}^{M} (KS(j) \times N(j)) \qquad \cdots (1)$$

[0065]Since the denominator of the right-hand side in a formula (1) is a value peculiar to an input string, it can be placed with ⸱A.⸱Generally the input string ⸱Okubo, Urawa⸱, The direction of a possibility of being a proper noun ⸱Urawa⸱, and a suffix ⸱city⸱ and a proper noun ⸱Okubo⸱ is judged to be higher than a possibility of being a proper noun ⸱Urawa⸱, a general noun ⸱city size⸱, and a proper noun ⸱Kubo⸱, and the size relation of both analysis likelihood serves as KS(1) >KS (2). Therefore, former one becomes high by ⸱Urawa⸱, probability KS(1) / A uttered, and ⸱Kubo⸱, and probability KS(2) / A uttered.

[0066]It can ask for the utterance probability of the utterance unit acquired from each analytical candidate based on each analytical candidate's analysis likelihood KS as mentioned above.

[0067]Next, the case where it asks for the above-mentioned utterance probability from the reading likelihood YS obtained by the above-mentioned reading grant part 13 is explained. The case of the above-mentioned input string ⸱Okubo, Urawa⸱ is raised to an example, and is explained. The above-mentioned analysis likelihood to all the given analytical candidates by the reading grant part 13. the 1st analytical candidate -- ⸱--- Urawa / city (**) -- suppose that the reading likelihood of the analytical candidate ⸱Urawa / city size (it is **** -- it is) / Kubo⸱ whose reading likelihood of the analytical candidate ⸱Urawa / city size (order) / Kubo⸱ whose reading likelihood of /Okubo (** to set)⸱ is YS (1) and the 2nd is YS (2) and the 3rd was calculated with YS (3).

[0068]The number of the utterance units acquired combining the composition word contained in the i-th analytical candidate is made into N(i). However, when the same utterance unit is included in two or more analytical candidates, suppose that the utterance unit of an analytical candidate with the highest analysis likelihood is counted.

[0069]The total combination of the composition word in one analytical candidate shall be generated in equivalent probability. If it does so, the utterance probability P2 of the utterance unit w in the i-th analytical candidate (w) can be expressed by a formula (2). However, M in a formula (2) is the number of analytical candidates.

$$P2(w) = YS(i) / \sum_{j=1}^{M} (YS(j) \times N(j)) \qquad \cdots (2)$$

[0070]Since the denominator of the right-hand side in a formula (2) is a value peculiar to an input string, it can be placed with ⸱B.⸱The size relation of each reading likelihood presupposes that it is YS(1) >YS(2) >YS (3). that is right -- then, ⸱reverse side -- ****** -- former one becomes high by **⸱ to set, probability YS(1) / B uttered, and ⸱reverse side ***********⸱, and probability YS(2) / B uttered.

[0071]It can ask for the utterance probability of the utterance unit acquired from each analytical candidate based on each analytical candidate's reading likelihood YS as mentioned above.

[0072]Next, the case where it asks for the above-mentioned utterance probability from the

above-mentioned word order of appearance is explained. Here, function h(i) which makes the word order of appearance i a variable is defined. Function h(i) is a function in which a value decreases as the word order of appearance i increases based on the rule of thumb that the probability of becoming an utterance unit as the word in the head part of the character string notation is high.

[0073]for example, the character string indicated to the program name of electronic program data -- "-- Mito Komon/53 -- order social reform trips and Shinagawa" are explained to an example. The word order of appearance makes N(i) the number of the utterance unit which contains at the head the word whose number is i.

[0074]All the utterance units which contain the same word at the head shall be generated in equivalent probability. namely, the utterance unit "Mito" which starts with a word "Mito" and "Mito Komon" -- "-- all the utterance probability of Mito Komon 53rd" turns into same probability. If it does so, the utterance probability P3 of the utterance unit w (w) that the word order of appearance contains at the head the word whose number is i can be expressed by a formula (3). However, M in a formula (3) is a word appearance order ordinal number (the number of words).

$$P\,3(w) = h(i) / \sum_{j=1}^{M} \left(h(j) \times N(j) \right) \qquad \cdots (3)$$

[0075]Since the denominator of the right-hand side in a formula (3) is a value peculiar to an analytical candidate, it can be placed with "C."the above-mentioned character string -- "-- Mito Komon/53 -- the result of the text analysis by the text analyzing part 12 to order social reform trips and Shinagawa" -- the word order of appearance i= 1 -- "Mito" and i= 2, as for "Komon" and i= 3, i= "the 53rd 4" are a "social reform", i= 5 is "traveled", and i= 6 becomes "Shinagawa."Based on the definition of function h(i), the word order of appearance i= 1 and the size relation of the value of function h(i) in 3 and 6 are set to h(1) >h(3) >h (6). Therefore, about an utterance unit "Mito Komon", "a 53rd social reform trip", and "Shinagawa", the utterance unit "a 53rd social reform trip" consists of an utterance unit "Shinagawa", and, in utterance probability, the utterance unit "Mito Komon" becomes high from an utterance unit "a 53rd social reform trip."therefore -- "-- I will see -- ****", and probability h(1) / C uttered -- "-- Mr. ****** -- next", and probability h(3) / C uttered -- "-- probability h(6) / C which carries out and is uttered with ******" -- "-- I will see -- the utterance probability of **** becomes the highest.

[0076]It can ask for the utterance probability of the utterance unit acquired based on the word order of appearance of each word which constitutes an input string as mentioned above.

[0077]Next, the case where it asks for the above-mentioned utterance probability from the above-mentioned number of Maura is explained. The utterance probability of an utterance unit becomes so low that the number of Maura of an utterance unit becomes large with a peak of a certain number of Maura. On the contrary, it becomes low even if the number of Maura

becomes small. Then, function m(i) showing the relation of the number i of Maura of an utterance unit and utterance probability that a key map is shown in <u>drawing 10</u> is defined.

[0078]The character string "murderous intent of the suspense masterpiece theater Tsugaru Tappi cape style" indicated to the program name of electronic program data is explained to an example. When uttering this program name, it is thought that "the murderous intent of the suspense masterpiece theater Tsugaru Tappi cape style" and the cases uttered as it is are rare, and their a possibility of being uttered in utterance units, such as a "suspense masterpiece theater", "suspense" or "murderous intent of the Tsugaru Tappi cape style", is high. At a "masterpiece" and a "theater", it is thought that there is too few Maura and utterance probability is low.

[0079]All the utterance units which present the same number of Maura shall be generated in equivalent probability. The number of Maura makes the number of the utterance unit which is i N(i). If it does so, the utterance probability P4 of the utterance unit w (w) that the number of Maura is i can be expressed by a formula (4). However, M in a formula (4) is the number of maximum Maura.

$$P\,4(w) = m(i) / \sum_{j=1}^{M} (m(j) \times N(j)) \qquad \cdots (4)$$

[0080]Since the denominator of the right-hand side in a formula (4) is a value peculiar to an analytical candidate, it can be placed with "D." As for the number of Maura of an utterance unit "murderous intent of the suspense masterpiece theater Tsugaru Tappi cape style", the number of Maura of 24 and an utterance unit "suspense theater" is set to 9 as a result of the text analysis by the text analyzing part 12 to the above-mentioned character string "murderous intent of the suspense masterpiece theater Tsugaru Tappi cape style." Based on the definition of function m(i), the size relation of i= 24 Maura and the value of function m(i) in 9 is set to m (9) >m (24). Therefore, latter one becomes high by "****************** to put is ****** of ******** Misa ******", probability m(24) / D uttered, and "**************** to put", and probability m(9) / D uttered.

[0081]It can ask for the utterance probability of the utterance unit based on the number of Maura of the utterance unit acquired from an input string as mentioned above.

[0082]Next, the case where it asks for the above-mentioned utterance probability from the above-mentioned word frequency of occurrence is explained. Here, function f(i) which makes a variable appearance frequency i in the character-string-information group which changes by all the character string information inputted into the text analyzing part 12 one by one is defined. This function f(i) is a function in which a value decreases as the character in which the word with much number of times which appears in the above-mentioned character-string-information group cannot turn into a word for distinguishing from other words easily is used and the appearance frequency i in the above-mentioned character-string-information group increases.

[0083]For example, the title ·Ministry of Finance homepage· of a homepage is explained to an example. Since a word ·homepage·, the word ·Welcome!·, etc. appear also in the title of other homepages frequently, the utterance probability of the utterance unit which contains these words as a recognized vocabulary for the call of a homepage becomes low.

[0084]All the utterance units containing the word which presents the same appearance frequency into the above-mentioned character-string-information group shall be generated in equivalent probability. The appearance frequency in the inside of the above-mentioned character-string-information group makes N(i) the number of the utterance unit containing the word which is i. If it does so, the utterance probability P5 of the utterance unit w concerned (w) that the appearance frequency of the word whose appearance frequency in the inside of the above-mentioned character-string-information group is the minimum among the words which constitute the utterance unit w is i can be expressed by a formula (5). However, M in a formula (5) is the maximum appearance frequency.

$$\mathrm{P}\ 5(\mathrm{w}) = \mathrm{f(i)} / \sum_{j=1}^{M} (\mathrm{f(j)} \times \mathrm{N(j)}) \qquad \cdots (5)$$

[0085]Since the denominator of the right-hand side in a formula (5) is a value peculiar to the above-mentioned character-string-information group, it can be placed with ·E.·While the utterance unit which contains the word the ·Ministry of Finance· in the above-mentioned character-string-information group is included once for example, supposing the utterance unit containing the word a ·homepage· is included 5 times, The size relation of the word frequency of occurrence i= 1 based on the definition of function f(i) and the value of function f(i) in 5 is set to f(1) >f (5). Therefore, former one becomes high by ·******** to set·, probability f(1) / E uttered, and ·**-****-**·, and probability f(5) / E uttered.

[0086]It can ask for the utterance probability of an utterance unit based on the word-as mentioned above frequency of occurrence in the above-mentioned character-string-information group.

[0087]When there is a title of five homepages as the above-mentioned character-string-information group, when the 1st title is inputted, specifically, it asks for the utterance probability P5 of the utterance unit W (w) first based on word frequency-of-occurrence i_1 in the character string of the title concerned, for example. Next, when the 2nd title is inputted, it is related with the same utterance unit W, and the utterance probability P5 (w) is recalculated based on word frequency-of-occurrence i_2 in the 1st character string of a title and the 2nd title. When the same operation is repeated below and the 5th title is finally inputted, It is related with the same utterance unit W, the utterance probability P5 (w) is recalculated based on word frequency-of-occurrence i_t in the 1st - all the 5th character strings of a title, and the utterance probability P5 of the final utterance unit W (w) is obtained.

[0088]Finally, the case where it asks for the above-mentioned utterance probability from the above-mentioned key word dictionary collation is explained. In this case, the key word dictionary which registered the keyword to which the probability value was given beforehand is drawn up. For example, a high probability value is given to a word "news" and it registers with the key word dictionary. On the contrary, a word "program" and the word "homepage" may be redundant, and there may be, or there may not be, and give and register the low probability value. [any] A default probability value is given to the word which is not registered into a key word dictionary.

[0089]By carrying out like this, former one becomes high with "****-**", the probability uttered, and "********" and the probability uttered. Here, giving the probability value "0" and registering with a key word dictionary plays a role equivalent to deleting from a recognized vocabulary.

[0090]Based on key word dictionary collation, it can ask for the utterance probability of the utterance unit as mentioned above.

[0091]It can be considered as the utterance probability of an utterance unit combining the utterance probability for which it asked using any [of six kinds of items of the "analysis likelihood", "reading likelihood", the "word order of appearance", the "number of Maura", the "word frequency of occurrence", and "key word dictionary collation" which were described above] one. As an example, it can ask for the utterance probability of an utterance unit by a formula like an equation (6). that is, If utterance probability which asked for the utterance probability which asked for the utterance probability which asked for the utterance probability which asked for the utterance probability which asked for the utterance probability for which it asked using analysis likelihood using P1 reading likelihood using P2 word order of appearance using the number of P3 Maura using P4 word frequency of occurrence using P5 key-word-dictionary collation is set to P6,

$$P = \sum_{i=1}^{6} m_i \cdot P_i \qquad \cdots (6)$$

It is here and is mi:weighting factor. [0092]As shown in <u>drawing 9</u>, utterance probability WS computed for every utterance unit as mentioned above is given to the recognized vocabulary (utterance unit) created by the lexical preparing part 15 like a 1st embodiment of the above, and is registered into the vocabulary storage section 17.

[0093]As mentioned above, in this embodiment, the text analyzing part 12 gives and outputs the analysis likelihood KS to all the division candidates, when two or more division candidates exist. When two or more reading exists, the reading grant part 13 is read to all the reading, and gives and outputs the likelihood YS. The utterance probability of the utterance unit acquired by the utterance probability calculation part 16 using at least one of the above-mentioned analysis likelihood KS and the reading likelihood YS, the word order of appearance, the number of Maura, the word frequency of occurrence, and the key word dictionary collation is computed. And he gives obtained utterance probability WS to a recognized vocabulary, and is trying to

register with the vocabulary storage section 17.

[0094]Therefore, according to this embodiment, utterance probability of the recognized vocabulary in the incorrect analysis in the recognized vocabulary registered into the dictionary for speech recognition drawn up by 1st embodiment of the above or the recognized vocabulary which is not uttered actually can be made low, The dictionary for speech recognition which can acquire high recognition precision can be drawn up.

[0095]Although the storing position in particular of the above-mentioned function h(i), m(i), f(i), and a key word dictionary is not limited, it shall have stored in the internal memory of the probability-of-occurrence calculation part 16, for example. In order to explain simply, he constitutes the utterance probability calculation part 16 in another block in the lexical preparing part 15, and is trying to compute utterance probability about the utterance unit acquired by the lexical preparing part 15. However, an utterance probability calculation part is constituted in the same block as a lexical preparing part, and although utterance probability calculation operation is incorporated after the above-mentioned step S4 in the flow chart of drawing 2, it does not interfere.

[0096]The <3rd embodiment> book embodiment is related with the voice recognition equipment carrying the dictionary for speech recognition drawn up by the dictionary preparation device for speech recognition in a 2nd embodiment of the above. Drawing 11 is a block diagram of the voice recognition equipment carrying the dictionary preparation device for speech recognition shown in drawing 6. The text analyzing part 23, the reading grant part 24, the analysis dictionary memory 25, the lexical preparing part 26, the utterance probability calculation part 27, and the vocabulary storage section 28 which constitute the dictionary preparation device 21 for speech recognition, It has the same composition as the above-mentioned text analyzing part 12 in a 2nd embodiment of the above, the reading grant part 13, the analysis dictionary memory 14, the lexical preparing part 15, the utterance probability calculation part 16, and the vocabulary storage section 17. And as shown in drawing 9, the recognized vocabulary to which utterance probability was given is stored in the vocabulary storage section 28.

[0097]On the other hand, the voice recognition equipment 22 comprises the acoustic analysis section 29, the likelihood computing part 30, the acoustics model storing part 31, and the collating part 32, and recognizes the sound inputted into the microphone using the recognized vocabulary information (dictionary for speech recognition) stored in the vocabulary storage section 28.

[0098]The above-mentioned acoustic analysis section 29 changes into a digital waveform the audio analog-spectrum form inputted from the microphone, and frequency analysis is conducted to every [about / 20msec-40msec] short time interval (frame), and it changes it into the vector system sequence of the parameter showing a spectrum. LPC (linear predictive

coding) mel cepstrum etc. are used for frequency analysis.

[0099]The above-mentioned likelihood computing part 30 calculates the likelihood of the acoustic model for every phonemes, such as HMM (Hidden Markov Model) stored in the acoustics model storing part 31, using the parameter vector of the voice inputting from the acoustic analysis section 29. In this way, it asks for the likelihood of each phoneme. The collating part 32 performs collation with the likelihood of each called-for phoneme, and all the recognized vocabularies registered into the vocabulary storage section 28, and computes the score of all the recognized vocabularies. In that case, the score of each recognized vocabulary which can be set is computed combining the sound likelihood which applied and obtained the likelihood of each phoneme for which the phoneme series of the recognized vocabulary concerned was asked in the likelihood computing part 30, and the language likelihood which is the utterance probability which is given to the recognized vocabulary concerned and memorized by the vocabulary storage section 28. for example, if sound likelihood of the recognized vocabulary W is set to p (W) and language likelihood is set to q (W), it will ask for the score score of the recognized vocabulary W (W) by score(W) = alpha-p (W)+ beta-q (W) however alpha, and beta:constant.

[0100]In this way, a score is computed about all the recognized vocabularies registered into the above-mentioned vocabulary storage section 28, and the recognition candidate who becomes by the recognized vocabulary which presents the score beyond a higher rank predetermined value is outputted as a recognition result.

[0101]As mentioned above, in this embodiment, voice recognition equipment has the vocabulary storage section 28 in which the dictionary for speech recognition drawn up by the dictionary preparation device for speech recognition of a 2nd embodiment of the above was stored. And the collating part 32 performs collation with the likelihood of each phoneme called for in the likelihood computing part 30, and the phoneme series of all the recognized vocabularies registered into the vocabulary storage section 28, He is trying to compute the score of all the recognized vocabularies combining the sound likelihood based on the likelihood of a phoneme, and the language likelihood which is the above-mentioned utterance probability.

[0102]Therefore, according to this embodiment, like the case of a 1st embodiment of the above, When a sound performs call of a homepage, and transfer in channel of television, even if it does not utter thoroughly the homepage title or program name which were registered, the call of a homepage and the change of a channel can be performed. The probability by which erroneous recognition is carried out to the recognized vocabulary obtained in the incorrect analysis of the above-mentioned input string information or the recognized vocabulary which is not uttered actually can be reduced by making the high-scoring recognized vocabulary based on the above-mentioned utterance probability (language likelihood) into a speech recognition

result in that case. Therefore, the voice recognition equipment which has high recognition performance is realizable.

[0103]In the above-mentioned embodiment, the above-mentioned voice recognition equipment 22 carries the dictionary preparation device 21 for speech recognition. However, the vocabulary storage section 28 by which the voice recognition equipment of this invention was created at least with the above-mentioned dictionary preparation device for speech recognition should just be carried, In the voice recognition equipment 22, even if it forms independently the text analyzing part 23, the reading grant part 24, the analysis dictionary memory 25, the lexical preparing part 26, and the utterance probability calculation part 27, they do not interfere.

[0104]The <4th embodiment> book embodiment is related with other examples of the voice recognition equipment carrying the dictionary for speech recognition drawn up by the dictionary preparation device for speech recognition in a 2nd embodiment of the above, and extracts from contents the character string information inputted into the above-mentioned dictionary preparation device for speech recognition.

[0105]Drawing 12 is a block diagram of the voice recognition equipment carrying the dictionary preparation device for speech recognition shown in drawing 6. The dictionary preparation device 41 for speech recognition. The text analyzing part 43 to constitute, the reading grant part 44, the analysis dictionary memory 45, the lexical preparing part 46, the utterance probability calculation part 47 and the vocabulary storage section 48, the acoustic analysis section 49 that constitutes the voice recognition equipment 42, the likelihood computing part 50, the acoustics model storing part 51, and the collating part 52, It has the same composition as the text analyzing part 23 in a 3rd embodiment of the above, the reading grant part 24, the analysis dictionary memory 25, the lexical preparing part 26, the utterance probability calculation part 27, the vocabulary storage section 28, the acoustic analysis section 29, the likelihood computing part 30, the acoustics model storing part 31, and the collating part 32. And as shown in drawing 9, the recognized vocabulary to which utterance probability was given is stored in the vocabulary storage section 48.

[0106]The incorporation part 53 incorporates contents including the character string information from the outside. The above-mentioned contents may also receive and incorporate with a receiver the information transmitted by broadcast, and, The information distributed on the Internet may also be incorporated via a communication network, and you may also incorporate from media which record data fixed, such as a magneto-optical disc, magnetic tape, a hard disk, and IC (integrated circuit) card. The character-string-information extraction part 54 extracts the character string information used for creation of the dictionary for speech recognition out of the contents incorporated in the incorporation part 53. And the extracted character string information is sent out to the text analyzing part 43 of the dictionary preparation device 41 for speech recognition.

[0107]When the tag information included in the above-mentioned contents is what kind of tag information, the extraction condition whether to extract character string information is stored in the extraction condition storage 55. And the character-string-information extraction part 54 extracts the character string information used for creation of the above-mentioned dictionary for speech recognition with reference to the above-mentioned extraction condition stored in the extraction condition storage 55.

[0108]Drawing 13 is a flow chart of the character-string-information extracting processing operation performed by the above-mentioned incorporation part 53 and the character-string-information extraction part 54. Hereafter, character-string-information extraction operation is explained according to drawing 13. The above-mentioned contents are incorporated by the above-mentioned incorporation part 53 at Step S21. Hereafter, it shifts to processing by the character-string-information extraction part 54. At Step S22, the initial character of the contents which were incorporated as for the account of the upper is read. At Step S23, it is distinguished whether the character by which reading appearance was carried out [above-mentioned] is empty (that is, contents the last). As a result, if it is the last, character-string-information extracting processing operation will be ended. On the other hand, if it is not the last, it will progress to Step S24. It is distinguished at Step S24 whether the received character is tag information. As a result, if it is not tag information, it will progress to Step S25. On the other hand, if it is tag information, it will progress to Step S26. At Step S25, after the following character in the above-mentioned contents is read, it returns to Step S23 and shifts to processing of the following character. It is distinguished whether the extraction condition of the extraction condition storage 55 is referred to, and the above-mentioned extraction condition is fulfilled by Step S26. As a result, in filling, while progressing to Step S27, in not filling, it returns to the above-mentioned step S25, and shifts to processing of the following character. At Step S27, the character string which fulfills the above-mentioned extraction condition is extracted, and it is sent out to the text analyzing part 43. It returns to the above-mentioned step S25, and such the back shifts to processing of the following character. Hereafter, processing of Step S23 - Step S27 is repeated, and in the above-mentioned step S23, contents end ******** and character-string-information extracting processing operation as it is the last.

[0109]For example, the above-mentioned contents are HTML (HyperText Markup Language) files, When an extraction condition "<title> tag exists in the extraction condition storage 55, supposing it is stored that the character string surrounded by <title> and </title> is extracted", If description top page </title> of the <title> Prime Minister's official residence Becoming is in the inputted contents (HTML file), the character string "top page of the Prime Minister's official residence" will be extracted.

[0110]In this way, by performing dictionary creation processing which is illustrated to drawing 2 with the dictionary preparation device 41 for speech recognition to the character string

automatically extracted from contents including character string information by the above-mentioned character-string-information extraction part 54, The optimal dictionary for speech recognition for the homepage call with a sound is drawn up and registered by the vocabulary storage section 48.Therefore, even if it utters thoroughly the title "top page of the Prime Minister's official residence" of the homepage of the Prime Minister's official residence to the voice recognition equipment 42, even if it utters selectively with the "Prime Minister's official residence" and "the top page of an official residence", it is recognized correctly, and the homepage of the Prime Minister's official residence can be called.

[0111]In that case, the utterance probability calculation part 47 is carried in the above-mentioned dictionary preparation device 41 for speech recognition, and the recognized vocabulary to which utterance probability was given is registered into the vocabulary storage section 48. Therefore, utterance probability of the recognized vocabulary generated as a result of the incorrect analysis to the character string extracted by the character-string-information extraction part 54 or a recognized vocabulary which is not uttered actually can be made low, and high recognition precision can be acquired.

[0112]In the voice recognition equipment shown in a 4th embodiment of the above, the <5th embodiment> book embodiment limits the contents from which the character string information inputted into the dictionary preparation device for speech recognition is extracted to web page information.

[0113]Drawing 14 is a block diagram of the voice recognition equipment carrying the dictionary preparation device for speech recognition shown in drawing 6. The dictionary preparation device 61 for speech recognition. The text analyzing part 63 to constitute, the reading grant part 64, the analysis dictionary memory 65, the lexical preparing part 66, the utterance probability calculation part 67 and the vocabulary storage section 68, the acoustic analysis section 69 that constitutes the voice recognition equipment 62, the likelihood computing part 70, the acoustics model storing part 71, and the collating part 72, It has the same composition as the text analyzing part 23 in a 3rd embodiment of the above, the reading grant part 24, the analysis dictionary memory 25, the lexical preparing part 26, the utterance probability calculation part 27, the vocabulary storage section 28, the acoustic analysis section 29, the likelihood computing part 30, the acoustics model storing part 31, and the collating part 32. And as shown in drawing 9, the recognized vocabulary to which utterance probability was given is stored in the vocabulary storage section 68.

[0114]The character-string-information extraction part 74 and the extraction condition storage 75 have the same composition as the character-string-information extraction part 54 and the extraction condition storage 55 in a 4th embodiment of the above. And the character string information extracted by the character-string-information extraction part 74 is outputted to the text analyzing part 63 of the dictionary preparation device 61 for speech recognition.

[0115]The web page information incorporation part 73 incorporates the web page information as the above-mentioned contents, and sends it out to the character-string-information extraction part 74 one by one from an initial character. Henceforth, the character string information which suits the extraction condition stored in the extraction condition storage 75 by the character-string-information extraction part 74 is extracted.

[0116]On the other hand, the control section 76 performs display control of a web page based on the speech recognition result by the above-mentioned voice recognition equipment 62. And the web page indicator 77 displays a web page according to directions of the control section 76.

[0117]Drawing 15 shows an example of the web page information incorporated by the above-mentioned web page information incorporation part 73. The information on a web page is described in languages, such as HTML. Supposing it is stored that and the character string surrounded by <title> and </title> is extracted when an extraction condition "<title> tag exists in the extraction condition storage 75", The character-string-information extraction part 74 performs extraction operation of the same character string information as the flow chart of drawing 13 fundamentally. That is, if the character string described by web page information is seen from the head and the <title> tag is found, the character string ·Ichiro Suzuki's homepage· inserted by <title> and </title> will be extracted.

[0118]In this way, by performing dictionary creation processing which is illustrated to drawing 2 with the dictionary preparation device 61 for speech recognition to the character string automatically extracted from web page information by the above-mentioned character-string-information extraction part 74, The dictionary for speech recognition in which the optimal recognized vocabulary with utterance probability for the homepage call with a sound was registered into the vocabulary storage section 68 is drawn up. To the utterance unit ·Suzuki· created from the above-mentioned character string ·Ichiro Suzuki's homepage· on that occasion, ·Ichiro·, ·Ichiro Suzuki·, ·Ichiro Suzuki's homepage·, and ·Ichiro's homepage.·URL (http://www.suzuki.xxx.jp etc.) of the homepage concerned is given, and he is trying to register with the vocabulary storage section 68, as shown in drawing 16.

[0119]Therefore, when uttered with ·Ichiro Suzuki· to the above-mentioned voice recognition equipment 62, Voice inputting is changed into a vector system sequence by the acoustic analysis section 69, and the likelihood of each phoneme is computed based on the above-mentioned vector system sequence in the likelihood computing part 70, Perform collation with the vocabulary of the vocabulary storage section 68 by the collating part 72, and it is recognized as the vocabulary ·Ichiro Suzuki·, and URL (http://www.suzuki.xxx.jp) given to the recognized vocabulary ·Ichiro Suzuki· is obtained.

[0120]If it does so, the above-mentioned control section 76 will access ·Ichiro Suzuki's homepage· based on obtained URL, will acquire the web page information on ·Ichiro Suzuki's

homepage", and will point to it and display the display of the web page concerned on the web page indicator 77.

[0121]That is, according to this embodiment, an utterance unit with utterance probability can be automatically generated from the information on a web page, and the suitable dictionary for speech recognition can be drawn up. That is, the dictionary for speech recognition as shown in Drawing I6 from a web page like drawing 15 can be drawn up automatically. Therefore, even if it utters the title "Ichiro Suzuki's homepage" of a web page thoroughly, Even if it utters with "Suzuki", "Ichiro Suzuki", and "Ichiro's homepage" selectively, it can recognize correctly, and the web page information on "Ichiro Suzuki's homepage" can be acquired, and it can display on the web page indicator 77.

[0122]When a homepage is registered into a browser by the bookmark of the above-mentioned browser, a favorite, etc., a title may be used as information which a user looks at and judges. Since the title in that case is not the character string information for being the information for seeing to the last, and uttering, an extremely long title may be given. Even in such a case, it becomes possible to call the homepage registered into the above-mentioned bookmark or a favorite by shorter utterance by drawing up the dictionary for speech recognition by the method which extracted and mentioned above the character string surrounded with the <title> tag.

[0123]In this embodiment, although the tag etc. into which a character font is changed besides the <title> tag although the case where the above-mentioned <title> tag was used was explained to the example store the extraction condition according to a use in the extraction condition storage 75, they are made. By doing so, it becomes possible to draw up the dictionary for speech recognition according to various uses.

[0124]It is also possible to store the changing condition of the above-mentioned utterance probability in the above-mentioned extraction condition storage 75. The changing condition of the utterance probability in that case is "choosing the high utterance unit of utterance probability from a title based on the notation of URL", for example. The alphabet notation about reading specifically included in "URL, It is that of making high utterance probability of the utterance unit containing the word (reading), if the coincidence condition of the notation (alphabet notation) of reading acquired from the result of the text analysis by the text analyzing part 63 and the reading grant part 64 and reading grant is investigated and there is a word in agreement."For example, if a title "homepage of Asahi Shimbun" and its URL "http://www.asahi.com" are taken for an example, The notation "asahi" of reading of the word "morning sun" obtained as a result of the alphabet notation "asahi" about reading included in URL, text analysis, and reading grant is the same. Therefore, the utterance probability of the utterance unit containing the word "morning sun" can be set up highly. In this way, it becomes possible to draw up simply the dictionary for speech recognition which makes possible high recognition precision about a title "homepage of Asahi Shimbun."

[0125]In the voice recognition equipment shown in a 4th embodiment of the above, the <6th embodiment> book embodiment limits the contents from which the character string information inputted into the dictionary preparation device for speech recognition is extracted to television program information.

[0126]Drawing 17 is a block diagram of the voice recognition equipment carrying the dictionary preparation device for speech recognition shown in drawing 6. The dictionary preparation device 81 for speech recognition. The text analyzing part 83 to constitute, the reading grant part 84, the analysis dictionary memory 85, the lexical preparing part 86, the utterance probability calculation part 87 and the vocabulary storage section 88, the acoustic analysis section 89 that constitutes the voice recognition equipment 82, the likelihood computing part 90, the acoustics model storing part 91, and the collating part 92, It has the same composition as the text analyzing part 23 in a 3rd embodiment of the above, the reading grant part 24, the analysis dictionary memory 25, the lexical preparing part 26, the utterance probability calculation part 27, the vocabulary storage section 28, the acoustic analysis section 29, the likelihood computing part 30, the acoustics model storing part 31, and the collating part 32. And as shown in drawing 9, the recognized vocabulary to which utterance probability was given is stored in the vocabulary storage section 88.

[0127]The character-string-information extraction part 94 and the extraction condition storage 95 have the same composition as the character-string-information extraction part 54 and the extraction condition storage 55 in a 4th embodiment of the above. And the character string information extracted by the character-string-information extraction part 94 is outputted to the text analyzing part 83 of the dictionary preparation device 81 for speech recognition.

[0128]The television-program-information incorporation part 93 incorporates the television program information as the above-mentioned contents, and sends it out to the character-string-information extraction part 94 one by one from an initial character. Incorporation of the above-mentioned television program information is performed by incorporating the electronic program data collected for one week [the part on 1] etc. Incorporation of this electronic program data may also be received and incorporated with a teletext receiver, and may also be incorporated via networks, such as the Internet. It does not interfere, although incorporated from archive media, such as a magneto-optical disc. In such the back, the character string information which suits the extraction condition stored in the extraction condition storage 95 by the character-string-information extraction part 94 is extracted.

[0129]On the other hand, the control section 96 controls the display of television, and recording and playback based on the speech recognition result by the above-mentioned voice recognition equipment 82. And the television display part 97 displays the image of television according to directions of the control section 96. The recording part 98 records a TV program according to directions of the control section 96. The regenerating section 99 plays the TV program

recorded by the recording part 98 according to directions of the control section 96.

[0130]In the television program listing, information, including the time of a program, a channel, a program name, etc., is indicated according to the decided form. Therefore, the dictionary for speech recognition can be drawn up like the above-mentioned 5 embodiments by extracting the character string of specific items, such as a program name, out of a television program listing.

[0131]Drawing 18 shows an example of the television program information incorporated by the above-mentioned television-program-information incorporation part 93. The extraction condition "a character string with a tag called a program name is extracted" shall be stored in the extraction condition storage 95. If it does so, the character-string-information extraction part 94 will perform extraction operation of the same character string information as the flow chart of drawing 13 fundamentally. namely, the character string which corresponds to the tag "program name" if the character string described by television program information is seen from the head and the tag "program name" is found -- "-- NHK news -- good morning, Japanese" is extracted.

[0132]In this way, by performing dictionary creation processing which is illustrated to drawing 2 with the dictionary preparation device 81 for speech recognition to the character string automatically extracted from television program information by the above-mentioned character-string-information extraction part 94, The dictionary for speech recognition in which the optimal recognized vocabulary with utterance probability for television program listing Shimesu with a sound was registered into the vocabulary storage section 88 is drawn up. that time -- the above-mentioned character string -- "-- NHK news good morning -- the utterance unit "NHK" created from Japanese", "news", and the "NHK news" -- "-- good morning, to Japanese" and "NHK news good morning."Channel information ("NHK synthesis" etc.), a day entry, time information, etc. of the program concerned are given, and he is trying to register with the vocabulary storage section 88, as shown in drawing 19.

[0133]Therefore, when uttered with "NHK news" to the above-mentioned voice recognition equipment 82, Voice inputting is changed into a vector system sequence by the acoustic analysis section 89, and the likelihood of each phoneme is computed based on the above-mentioned vector system sequence in the likelihood computing part 90, Perform collation with the vocabulary of the vocabulary storage section 88 by the collating part 92, and it is recognized as the vocabulary "NHK news", and the channel information (NHK synthesis), the day entry (May 5), and time information (5:00 to 8:15) which are given to the recognized vocabulary "NHK news" are acquired.

[0134]pointing to the display of television to the television display part 97 based on the channel information, the day entry, and time information from which that was right, then the above-mentioned control section 96 was obtained -- 5:00 a.m. on May 5 -- "-- NHK news -- good

morning, Japanese· is displayed.

[0135]That is, according to this embodiment, an utterance unit with utterance probability can be automatically generated from television program information, and the suitable dictionary for speech recognition can be drawn up. That is, the dictionary for speech recognition as shown in Drawing I9 from television program information as shown in drawing 18 can be drawn up automatically. therefore, the program name of a TV program -- ·-- NHK news -- good morning, even if it utters Japanese· thoroughly, partial -- ·NHK·· -- ·-- good morning -- recognizing correctly, even if it utters with Japanese· and ·NHK news good morning· -- a TV program -- ·-- NHK news -- good morning, the channel information, the day entry, and time information of Japanese· can be acquired, and it can display on the television display part 97.

[0136]In this embodiment, although the case where the above-mentioned program name tag was used was raised to the example, and was explained, and various extraction conditions are stored in the extraction condition storage 95 according to the use, it can do. For example, when reserving recording, since the previous program should become a recognition object rather than a present date and time, it is also possible to store the extraction condition of ·the date extracts subsequent (or time, current time or subsequent ones) program names today· in the extraction condition storage 95.

[0137]The <7th embodiment> book embodiment is related with the system which controls information home appliance apparatus by a voice remote control about the example of use of a 6th embodiment of the above. This information home appliance apparatus voice remote control system has composition as shown in drawing 20. The voice remote control 101 comprises the microphone 102, the speaker 103, and the remote-control part 104. And if a sound is inputted into the microphone 102, voice inputting will be sent out to the voice recognition equipment 106 by the remote-control part 104 via the communication line 105. The speaker 103 outputs with a sound the recognition result etc. which are sent out from the voice recognition equipment 106 via the communication line 105, and is used for the check of a recognition result, etc. Infrared rays etc. are used for the communication line 105, for example.

[0138]The voice recognition equipment 106 has the composition which deleted the control section 96, the television display part 97, the recording part 98, and the regenerating section 99 from the composition of the voice recognition equipment carrying the dictionary preparation device for speech recognition shown in drawing 17. And the sound inputted into the acoustic analysis section (not shown) from the communication line 105 is recognized, and a recognition result is sent out to the information home appliance apparatus 108 via the communication line 107. In that case, the television program information inputted into the television-program-information incorporation part (not shown) of the dictionary preparation device for speech recognition is inputted from the information home appliance apparatus 108 via the communication line 107. Like ****, the communication line 107 is a network which connects the

voice recognition equipment 106 and the information home appliance apparatus 108, and are networks, such as LAN (local area network) of a cable or radio. The voice recognition equipment 106 can be considered to be a personal computer, and, specifically, the information home appliance apparatus 108 can be considered to be television. In this example, although it has composition which connects the voice recognition equipment 106 and the information home appliance apparatus 108 by the communication line 107, it does not matter at all although the voice recognition equipment 106 is incorporated in the information home appliance apparatus 108.

[0139]The appliance control part 109 in which the above-mentioned information home appliance apparatus 108 performs various control based on the information from the above-mentioned communication line 107, The main information memory 110 which memorizes the main information used as the print-out of the apparatus concerned, such as an image and music, The main information outputting part 112 which has the sub information memory 111 which memorizes the sub information in connection with the above-mentioned main information, such as a program name of an image and a musical title name, a display, a speaker, etc., and outputs the above-mentioned main information is carried.

[0140]The communication line 113 is a network which connects the above-mentioned information home appliance apparatus 108 and the outdoor offer-of-information center 114, and are a network and digital broadcasting networks, such as a telephone line and a cable TV circuit. The offer-of-information center 114 is equivalent to a provider or a broadcasting station, and provides main information via the communication line 113 to the information home appliance apparatus 108. The main information memory 115 which memorizes the main information which should send out this offer-of-information center 114 to the information home appliance apparatus 108, It has the control section 117 which controls the memory of a variety of information to the sub information memory 116 which memorizes the sub information in connection with main information, and each memory 115,116, read-out, sending out in the communication line 113 of read-out information, etc.

[0141]In the information home appliance apparatus voice remote control system which has the above-mentioned composition, with the sound inputted into the voice remote control (wireless microphone) 101, main information can be chosen and it can reproduce by the nearby information home appliance apparatus 108. Operation of this information home appliance apparatus voice remote control system is concretely explained to an example for the case where utter with ·NHK news· toward the voice remote control 101 at a home, and ·NHK news· is made to output to TV footage hereafter.

[0142]The television program information as the television imagery information and sub information as main information which the above-mentioned offer-of-information center 114 provides is incorporated into the information home appliance apparatus 108 via the

communication line 113, and television program information is stored in the sub information memory 111 by the appliance control part 109. And the appliance control part 109 sends out the memory content (television program information) of the sub information memory 111 to the voice recognition equipment 106 via the communication line 107.

[0143]As the above-mentioned voice recognition equipment 106 incorporates the received television program information into the television-program-information incorporation part of the above-mentioned dictionary preparation device for speech recognition and stated it by explanation of drawing 17, it performs extraction of character string information, and creation of the dictionary for speech recognition. As a result, utterance probability, and channel information, a day entry and time information are given to the drawn-up dictionary for speech recognition, and the recognized vocabulary for TV program specification as shown in drawing 19 is registered into it.

[0144]If it is uttered with ·NHK news· to the microphone 102 of the above-mentioned voice remote control 101 in this state, The inputted sound is incorporated into the acoustic analysis section of the voice recognition equipment 106 via the communication line 105, and is made above, Conversion in a vector system sequence, likelihood calculation of each phoneme, and collation with the above-mentioned dictionary for speech recognition are performed, and it is recognized as the vocabulary ·NHK news·, and the channel information, the day entry, and time information which are given to the recognized vocabulary ·NHK news· are acquired. And it transmits to the information home appliance apparatus 108 via the communication line 107 by making these information into a command.

[0145]So then, the appliance control part 109 of the above-mentioned information home appliance apparatus 108, If the received command is compared with the contents of the sub information memory 111, and is interpreted and the date and time by the above-mentioned day entry and time information come, the video information of the NHK synthesis by the above-mentioned channel information will be outputted to the main information outputting part 112. When the received command is a recording command, the above-mentioned video information is stored in the main information memory 110. When the received command is a reproduction command, the above-mentioned video information stored in the main information memory 110 is read, and it is outputted to the main information outputting part 112.

[0146]Thus, according to this embodiment, it is made to perform voice input to the acoustic analysis section of the above-mentioned voice recognition equipment 106 via the communication lines 105, such as infrared rays, from the voice remote control 101. therefore, the operativity of an information home appliance apparatus voice control system -- many -- it can improve.

[0147]The above-mentioned voice recognition equipment 106 is voice recognition equipment shown in drawing 17, and the information home appliance apparatus voice remote control

system shown in drawing 20 explains to an example the case where television program information is inputted as sub information. However, this invention is not limited to this, may use the voice recognition equipment 106 as the voice recognition equipment shown in drawing 14, and may use web page information as sub information. Or the voice recognition equipment 106 may be used as the voice recognition equipment shown in drawing 12, and general contents information may be used as sub information.

[0148]According to <8th embodiment> above-mentioned each embodiment, the utterance unit which changes with several connection words from which the combination of the word obtained as a result of text analysis differs is registered into the dictionary for speech recognition as it is as a recognized vocabulary, without taking into consideration in any way to the acoustical similarity between each utterance unit. Therefore, when the title of a homepage has ·a top page of the Prime Minister's official residence·, and ·the opinion of the Department of Justice·, erroneous recognition of the utterance ·prime minister --· may be carried out to ·opinion --·, and the homepage of ·the opinion of the Department of Justice· may be displayed. In such a case, this embodiment copes with it.

[0149]Drawing 21 is a block diagram of the voice recognition equipment carrying the dictionary preparation device for speech recognition shown in drawing 6. The dictionary preparation device 121 for speech recognition. The text analyzing part 123 to constitute, the reading grant part 124, the analysis dictionary memory 125, the utterance probability calculation part 127 and the vocabulary storage section 128, the acoustic analysis section 129 that constitutes the voice recognition equipment 122, the likelihood computing part 130, the acoustics model storing part 131, and the collating part 132, It has the same composition as the text analyzing part 23 in a 3rd embodiment of the above, the reading grant part 24, the analysis dictionary memory 25, the utterance probability calculation part 27, the vocabulary storage section 28, the acoustic analysis section 29, the likelihood computing part 30, the acoustics model storing part 31, and the collating part 32. And as shown in drawing 9, the recognized vocabulary to which utterance probability was given is stored in the vocabulary storage section 128.

[0150]The similarity calculating part 133 calculates the acoustical similarity of two arbitrary utterance units created by the lexical preparing part 126 like the lexical preparing part 26 in a 3rd embodiment of the above. As a result, it turns out that the utterance unit ·prime minister· and the utterance unit ·opinion· are acoustically similar from the utterance unit ·prime minister· and the utterance unit ·official residence·, for example. Although there are various methods in calculation of the above-mentioned similarity, it is realizable by calculating based on how many the phoneme which is in agreement, for example exists.

[0151]And the lexical preparing part 126 changes the utterance probability value of each utterance unit calculated in the above-mentioned utterance probability calculation part 127 using the calculation result of the similarity by the similar calculation part 133.

[0152]Here, suppose that the title of a homepage has "a top page of the Prime Minister's official residence", and "the opinion of the Department of Justice."By a series of processings depended by the lexical preparing part 126 from the text analyzing part 123, an utterance unit "prime minister", an "official residence", "the Department of Justice (carrying out way ******)", an "opinion", etc. are searched for with utterance probability. In that case, when the utterance probability of an utterance unit "prime minister" and an utterance unit "opinion" is compared, it is as a 2nd embodiment of the above having described that a value with former high one is given. However, since the utterance unit "opinion" has appropriate utterance probability, when it is registered into the vocabulary storage section 128 as it is and used as the dictionary for speech recognition, it has much possibility that erroneous recognition of the utterance "prime minister" may be carried out to an "opinion", and the homepage of "the opinion of the Department of Justice" may be displayed.

[0153]Then, the above-mentioned lexical preparing part 126 searches two acoustically similar utterance units based on the similarity by the similarity calculating part 133. And one utterance unit plays a central role in a homepage title, when the utterance unit of another side is not so, the utterance probability of a central utterance unit is raised further, and the utterance probability of the utterance unit of the direction which is not so is lowered further. In that case, it can be judged whether a role with an applicable central utterance unit is played by the size of the probability value given in the utterance probability calculation part 127. Into how [Mr.] an utterance probability value is changed in that case does not especially specify. For example, a changing amount may be set up according to the value of an utterance probability value, and a changing amount may be set up based on similarity. It is also possible to change into "0" the utterance probability value of an utterance unit which has not played the central role.

[0154]As mentioned above, in this embodiment, the similarity of two arbitrary utterance units generated by the lexical preparing part 126 is calculated by the similarity calculating part 133, When two acoustically similar utterance units exist, the lexical preparing part 126 is made to make low probability of the utterance unit which is not a central role between two acoustically similar utterance units. therefore -- according to this embodiment -- the above -- the time of uttering the utterance unit which plays a central role -- the above -- it can prevent that erroneous recognition is carried out to the utterance unit which has not played the central role. That is, according to this embodiment, the voice recognition equipment whose recognition performance is still higher can be built.

[0155]In <9th embodiment> above-mentioned each embodiment, the utterance unit generated from an address book, web page information, or television program information is registered into the vocabulary storage sections 6, 17, 28, 48, 68, and 88,128 as a vocabulary for recognition. Therefore, the vocabulary storage sections 6, 17, 28, 48, 68, and 88,128 show a high recognition rate, when the title and program name of an address or a homepage which

are registered are uttered. However, a user does not necessarily utter the address, title name, or program name which are always registered into the vocabulary storage sections 6, 17, 28, 48, 68, and 88,128. Therefore, the problem that recognition precision when the address, title name, and program name which are not registered are uttered falls extremely will arise. In such a case, this embodiment copes with it.

[0156]Drawing 22 is a block diagram of the voice recognition equipment in this embodiment. The acoustic analysis section 142, the likelihood computing part 143, the acoustics model storing part 144, and the 1st vocabulary storage section 146 which constitute the voice recognition equipment 141, It has the same composition as the acoustic analysis section 29, the likelihood computing part 30, the acoustics model storing part 31, and the vocabulary storage section 28 which constitute the voice recognition equipment 22 in a 3rd embodiment of the above. And as shown in drawing 16, the recognized vocabulary for homepage titles to which utterance probability was given shall be stored in the 1st vocabulary storage section 146.

[0157]The 2nd vocabulary storage section 147 is not created based on a specific text by each dictionary preparation device for speech recognition mentioned above, but is the dictionary for speech recognition in which the fixed general vocabulary was registered, and is compared by the collating part 145 like the 1st vocabulary storage section 146. That is, the 2nd vocabulary storage sections 147 differ in the point that the fixed general vocabulary is registered to the 1st vocabulary storage section 146 having registered only the vocabulary which the user chose.

[0158]The above-mentioned collating part 145 performs collation with the vocabulary of the 1st vocabulary storage section 146 and the 2nd vocabulary storage section 147, and when it is judged that the vocabulary registered into the 1st vocabulary storage section 146 is a recognized vocabulary, it outputs the sub information (URL) given to the vocabulary and vocabulary concerned as a recognition result. On the other hand, when it is judged that the vocabulary registered into the 2nd vocabulary storage section 147 is a recognized vocabulary, it sends out to the retrieval part 148 by making the vocabulary into a recognition result. So then, the character string information corresponding to the recognition result which the retrieval part 148 received from the collating part 145, In being the utterance unit generated from web page information to television program information when it was the utterance unit generated from an address book to web page information when the vocabulary registered into the 1st vocabulary storage section 146 was the utterance unit generated from the address book, it searches from television program information. The selecting part 149 is for a user to choose the lexical information and grant information which are registered into the 1st vocabulary storage section 146, and to register with the 1st vocabulary storage section 146 out of the character string information searched [above-mentioned] with the vocabulary (the above-mentioned recognition result) which the retrieval part 148 received.

[0159]Hereafter, the vocabulary registered into the 1st vocabulary storage section 146 of the above explains to an example the case where it is the utterance unit generated from web page information, about operation of the voice recognition equipment in this embodiment. In that case, in a real-intention voice recognition device, the web page according to a recognition result can be displayed by uttering the vocabulary registered into the 1st vocabulary storage section 146. The character string information (URL) about the pair and the recognized vocabulary concerned of a recognized vocabulary/reading is matched and registered into the 1st vocabulary storage section 146 at least. In that case, what it could be created by the method described by a 5th embodiment of the above, and the user inputted by himself may be sufficient as the recognized vocabulary registered. Even if it carries out any, it is the 1st vocabulary storage section 146 that limited to the web page which a user often looks at, and registered the recognized vocabulary about it.

[0160]By the way, the vocabulary which a user utters is not always in the 1st vocabulary storage section 146. Then, the 2nd vocabulary storage section 147 is formed so that a recognition result may be obtained, even when the vocabulary which is not in the 1st vocabulary storage section 146 is uttered. Not only the recognized vocabulary that the user registered but the arbitrary vocabularies generally used are memorized by this 2nd vocabulary storage section 147. It is extensive over the range in which the contents of registration of the 2nd vocabulary storage section 147 are general to some extent wide at immobilization to the contents of registration of the 1st vocabulary storage section 146 being the small number limited to the range which generally exists with variable.

[0161]The above-mentioned collating part 145 judges whether it compares with the contents of registration of the 1st vocabulary storage section 146 at the time of recognition, or it compares with the contents of registration of the 2nd vocabulary storage section 157. Do not limit, especially if attached to the method, and it compares with the 1st vocabulary storage section 146 first, for example, When recognition likelihood is not high enough, the method of performing collation with the 2nd vocabulary storage section 147 may be used, it may compare by both the 1st vocabulary storage section 146 and the 2nd vocabulary storage section 147, and the method of making the high recognition candidate of recognition likelihood a recognition result may be used.

[0162]Here, when the group of a recognized vocabulary / reading / utterance probability / URL as shown in drawing 16 at the 1st vocabulary storage section 146 of the above is registered, suppose that the user uttered with ·Sato.·In that case, since it reads to the 1st storage parts store 146 and ·satou· does not exist, utterance ·Sato· is not recognized. In that case, the vocabulary ·Sato· will be recognized by collation with the 2nd vocabulary storage section 147. In this way, the obtained recognition result ·Sato· is sent out to the retrieval part 148. And search of a web page is performed by the retrieval part 148 by a keyword ·Sato.·Search of this

web page is realizable by using the search engine which has spread on the Internet. A search engine is a program which discovers URL of the web page relevant to it from the given keyword. Generally URL discovered has more than one, and a search engine shows them to a user. And if URL of desired "Mr. Sato's homepage" is chosen by the user out of the search results by the retrieval part 148, the selecting part 149 will associate a recognized vocabulary "Sato/satou" and URL of "Mr. Sato's homepage", and will register them into the 1st vocabulary storage section 146 by him.

[0163]In this way, if a recognized vocabulary "Sato/satou" and URL of "Mr. Sato's homepage" are registered into the 1st vocabulary storage section 146 of the above, Without performing search and selection henceforth, if it utters with "Sato", the desired web page "Mr. Sato's homepage" can be seen promptly. Registration of the new recognized vocabulary to the 1st vocabulary storage section 146 can also be easily performed only by selection by utterance and the selecting part 149, without inputting a character.

[0164]As for the voice recognition equipment using the dictionary for speech recognition drawn up with the dictionary preparation device for speech recognition in each above-mentioned embodiment, it is effective to carry in personal digital assistant machines, such as a cellular phone and an electronic notebook. Namely, it is better for operativity to be based on utterance rather than key operation in such a personal digital assistant machine, when performing operator guidance. However, it is difficult to utter correctly [as decided beforehand] wording for performing operator guidance in a destination etc., and it still more difficult for a user to draw up the dictionary for speech recognition for coping with it in such a case.

[0165]According to the dictionary preparation device for speech recognition in each above-mentioned embodiment, the combination of all the division candidates, all the reading candidates, and all the connection words is taken into consideration from one input string, Since the vocabulary for recognition which becomes per two or more utterance is automatically generable, even if it utters the partial character strings of the character string set up beforehand, the dictionary for speech recognition which can be recognized correctly can be drawn up very easily. Therefore, it is dramatically effective as a voice synthesizer for personal digital assistant machines to carry the voice recognition equipment using such a dictionary for speech recognition.

[0166]In a place. The above-mentioned analysis means by the above-mentioned text analyzing part in each above-mentioned embodiment, the reading grant part, the lexical preparing part, the vocabulary storage section, the utterance probability calculation part, the incorporation part, the character-string-information extraction part, and a similarity calculating part, a reading grant means, a lexical preparing means, a lexical memory measure, an utterance probability calculating means, an incorporation means, The function as a character-string-information extraction means and a similarity calculation means is realized by the dictionary creation

processing program recorded on the program recording medium. The above-mentioned program recording media in the above-mentioned embodiment are program media which become by ROM (read only memory). Or they may be the program media equipped with and read to external auxiliary storage. The program reading means which reads a dictionary creation processing program from the above-mentioned program media in the case of which, It may have the composition which carries out direct access to the above-mentioned program media, and is read to them, and it may download in the program store area (not shown) established in RAM (random access memory), and may have the composition accessed and read to the above-mentioned program store area. The download program for downloading in the above-mentioned program store area of RAM from the above-mentioned program media shall be beforehand stored in the main frame.

[0167]With the above-mentioned program media, it is constituted disengageable the main part side here, Magnetic disks, such as a tape system of magnetic tape, a cassette tape, etc., a floppy (registered trademark) disk, and a hard disk, CD(compact disk)-ROM and MO (optical magnetism) disk, MD (mini disc), The disk system of optical discs, such as DVD (digital video disc), It is a medium including semiconductor memory systems, such as card systems, such as IC (integrated circuit) card and an optical card, a mask ROM, EPROM (ultraviolet-rays elimination type ROM), EEPROM (electric elimination type ROM), and a flash ROM, which supports a program fixed.

[0168]The voice recognition equipment in each above-mentioned embodiment, If it has the composition which is provided with a modem and contains the Internet and in which a communication network and connection are possible, even if the above-mentioned program media are media which support a program fluidly by download from a communication network, etc., they will not interfere. The download program for downloading from the above-mentioned communication network which can be set in that case shall be beforehand stored in the main frame. Or it shall be installed from another recording medium.

[0169]It is not limited only to a program and what is recorded on the above-mentioned recording medium can also record data.

[0170]

[Effect of the Invention]As mentioned above, so that clearly the dictionary preparation device for speech recognition of the 1st invention, Based on the reading candidate of all ************, for all the division candidates and reading grant means which were obtained from one character string information by the analysis means, by a lexical preparing means. Since one or more utterance units to which reading was given are generated and it registers by a lexical memory measure by making the generated utterance unit into a recognized vocabulary, the dictionary for speech recognition which makes an utterance unit with the possibility of utterance a recognized vocabulary from the given character string information is generable.

Therefore, even if a user utters which partial character strings in the character string set up beforehand, the dictionary for speech recognition for recognizing correctly can be drawn up by low cost.

[0171]The dictionary preparation device for speech recognition of an invention of the above 1st, By an utterance probability calculating means, calculate the utterance probability of each utterance unit generated [above-mentioned] using at least one of analysis likelihood, reading likelihood, the word order of appearance, the number of Maura, the word frequency of occurrence, and the key word dictionary collated results, and by the above-mentioned lexical preparing means. If the above-mentioned utterance probability is given and the above-mentioned lexical memory measure is made to memorize the recognized vocabulary which becomes in each above-mentioned utterance unit, Even if the recognized vocabulary generated as a result of the incorrect analysis by the above-mentioned analysis means and the recognized vocabulary which is not uttered actually are registered into the above-mentioned dictionary for speech recognition, the utterance probability of such an unnecessary recognized vocabulary is set up small, and the dictionary for speech recognition which can realize high recognition precision can be drawn up.

[0172]The dictionary preparation device for speech recognition of an invention of the above 1st, An incorporation means to incorporate contents including character string information, and the extraction condition storing means in which the extraction condition of character string information required for dictionary creation was stored, If it has a character-string-information extraction means to extract character string information from the above-mentioned contents with reference to the above-mentioned extraction condition, and to send out to the above-mentioned analysis means, the above-mentioned dictionary for speech recognition can be automatically drawn up from the above-mentioned contents information.

[0173]The dictionary preparation device for speech recognition of an invention of the above 1st, If the above-mentioned incorporation means is accomplished so that the information on a web page may be incorporated as the above-mentioned contents, By storing in the above-mentioned extraction condition storing means "when a <title> tag exists, the character string surrounded by <title> and </title> is extracted", for example, The character string "top page of the Prime Minister's official residence" is extracted from the top page </title> of the title "<title> Prime Minister's official residence of a web page", and the above-mentioned dictionary for speech recognition can be drawn up automatically.

[0174]The dictionary preparation device for speech recognition of an invention of the above 1st, If the above-mentioned picking ****** is accomplished so that the information on a TV program may be incorporated as the above-mentioned contents, the character string which corresponds to a tag "program name" by storing in the above-mentioned extraction condition storing means "a character string with a tag called a program name is extracted", for example -

- "-- NHK news -- good morning, Japanese" is extracted and the above-mentioned dictionary for speech recognition can be drawn up automatically.

[0175]The dictionary preparation device for speech recognition of an invention of the above 1st, Have a similarity calculation means which calculates the acoustical similarity between each utterance unit generated by the above-mentioned lexical preparing means, and by the above-mentioned lexical preparing means. If it accomplishes so that the utterance probability given to each above-mentioned recognized vocabulary may be changed according to the above-mentioned similarity, when the utterance unit "prime minister" and the utterance unit "opinion" are acoustically similar, For example, by lowering further the value of the utterance probability of the utterance unit "opinion" which is not so, while raising further the value of the utterance probability of an utterance unit "prime minister" of the value of the above-mentioned utterance probability being high, and having played the central role in the input string, It can prevent that erroneous recognition of the utterance "prime minister" which plays a central role is carried out to an "opinion."

[0176]The dictionary preparation method for speech recognition of the 2nd invention, The step which analyzes the inputted character string information, divides into a composition word, and outputs all the division candidates, The step which gives reading to each composition word divided [above-mentioned], and outputs all the reading candidates, Since it had the step which generates one or more utterance units to which reading was given as a recognized vocabulary, and the step which memorizes each recognized vocabulary generated [above-mentioned] as a dictionary for speech recognition based on the division candidate and reading candidate of all above, The dictionary for speech recognition which makes an utterance unit with the possibility of utterance a recognized vocabulary from the given character string information like the case of an invention of the above 1st is generable. Therefore, even if it utters which partial character strings in the character string set up beforehand, the dictionary for speech recognition for recognizing correctly can be drawn up.

[0177]Since the dictionary for speech recognition drawn up by the dictionary preparation device for speech recognition of the invention of the above 1st is used for the voice recognition equipment of the 3rd invention as a dictionary for collation, By using the dictionary for speech recognition which makes a recognized vocabulary an utterance unit with the possibility of utterance generated from the given character string information, even if it utters the partial character strings of the character string set up beforehand, it can recognize correctly.

[0178]The voice recognition equipment of an invention of the above 3rd with the dictionary preparation device for speech recognition which incorporates the information on the above-mentioned web page, using the drawn-up dictionary for speech recognition as the above-mentioned dictionary by a control means. If the display information of a web page displaying means is switched and controlled based on a recognition result and the web page according to

the above-mentioned recognition result is displayed, the web page according to an uttered content can be displayed correctly.

[0179]The voice recognition equipment of an invention of the above 3rd with the dictionary preparation device for speech recognition which incorporates the information on the TV program text-ized [above], using the drawn-up dictionary for speech recognition as the above-mentioned dictionary by a control means. If a television displaying means, a recording means, and a reproduction means are controlled based on a recognition result, according to an uttered content, change of a display channel, setting out of a recording condition, or reproduction of a picture recording program can be performed correctly.

[0180]If the voice recognition equipment of an invention of the above 3rd is provided with the auxiliary dictionary drawn up without being based on the dictionary preparation device for speech recognition of an invention of the above 1st and a collation means is made to perform collation with the above-mentioned dictionary and an auxiliary dictionary, Even when the vocabulary which is not registered into the above-mentioned dictionary is uttered, the vocabulary can be recognized correctly. When the recognized vocabulary of the above-mentioned auxiliary dictionary is chosen as a recognition result as a result of the above-mentioned collation, By a search means, out of the web page information relevant to the title of a web page, for example, it was inputted into the above-mentioned dictionary preparation device for speech recognition, search the character string applicable to the recognition result concerned, and by a selecting means. If the character string registered into the above-mentioned dictionary is chosen from two or more character strings searched [above-mentioned], by registering the character string into the above-mentioned dictionary, a recognized vocabulary can be increased and recognition speed can be improved.

[0181]Since the dictionary for speech recognition which the voice recognition equipment of the 4th invention carried the dictionary preparation device for speech recognition of the invention of the above 1st, and was drawn up by the above-mentioned dictionary preparation device for speech recognition is used as a dictionary for the above-mentioned collation, By inputting character string information into the dictionary preparation device for speech recognition carried [above-mentioned], even if it utters which partial character strings in the character string set up beforehand, the dictionary for speech recognition for recognizing correctly can be drawn up automatically. Therefore, high recognition precision can be acquired.

[0182]Since the personal digital assistant machine of the 5th invention carried the voice recognition equipment of the above 3rd and the 4th invention, even if it does not utter wording for performing operator guidance correctly as decided beforehand, in a destination etc., the voice call of a homepage, etc. can be performed correctly, for example.

[0183]Since the dictionary creation processing program as which a computer is operated as the analysis means in the 1st above-mentioned invention, a reading grant means, a lexical

preparing means, and a lexical memory measure is recorded, the program recording medium of the 6th invention, One or more utterance units to which reading was given as well as the case of an invention of the above 1st can be generated, and it can register as a recognized vocabulary. Therefore, even if it utters which partial character strings in the character string set up beforehand, the dictionary for speech recognition for recognizing correctly can be drawn up.

[Translation done.]